SRI International

A PRACTICAL NONMONOTONIC THEORY FOR REASONING ABOUT SPEECH ACTS

Technical Note 432

April 6, 1988

By: Douglas Appelt
Senior Computer Scientist

Kurt Konolige
Computer Scientist

Artificial Intelligence Center
Computer and Information Sciences Division

This paper to appear in *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics.*

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

The research was supported in part by a contract with the Nippon Telegraph and Telephone Corporation, in part by the Office of Naval Research under Contract N00014-85-C-0251, and in part under subcontract with Stanford University under Contract N00039-84-C-0211 with the Defense Advanced Research Projects Agency.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
06 APR 1988		**00-04-1988 to 00-04-1988**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Practical Nonmonotonic Theory for Reasoning About Speech Acts	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**10**	

A Practical Nonmonotonic Theory for Reasoning about Speech Acts

Douglas Appelt, Kurt Konolige
Artificial Intelligence Center and
Center for the Study of Language and Information
SRI International
Menlo Park, California

Abstract

A prerequisite to a theory of the way agents understand speech acts is a theory of how their beliefs and intentions are revised as a consequence of events. This process of attitude revision is an interesting domain for the application of nonmonotonic reasoning because speech acts have a conventional aspect that is readily represented by defaults, but that interacts with an agent's beliefs and intentions in many complex ways that may override the defaults. Perrault has developed a theory of speech acts, based on Rieter's default logic, that captures the conventional aspect; it does not, however, adequately account for certain easily observed facts about attitude revision resulting from speech acts. A natural theory of attitude revision seems to require a method of stating preferences among competing defaults. We present here a speech act theory, formalized in hierarchic autoepistemic logic (a refinement of Moore's autoepistemic logic), in which revision of both the speaker's and hearer's attitudes can be adequately described. As a collateral benefit, efficient automatic reasoning methods for the formalism exist. The theory has been implemented and is now being employed by an utterance-planning system.

1 Introduction

The general idea of utterance planning has been at the focus of much NL processing research for the last ten years. The central thesis of this approach is that utterances are actions that are planned to satisfy particular speaker goals. This has led researchers to formalize speech acts in a way that would permit them to be used as operators in a planning system [1,2]. The central problem in formalizing speech acts is to correctly capture the pertinent facts about the revision of the speaker's and hearer's attitudes that ensues as a consequence of the act. This turns out to be quite difficult because the results of the attitude revision are highly conditional upon the context of the utterance.

To consider just a small number of the contingencies that may arise, consider a speaker S uttering a declarative sentence with propositional content P to hearer H. One is inclined to say that, if H believes S is sincere, H will believe P. However, if H believes $\neg P$ initially, he may not be convinced, even if he thinks S is sincere. On the other hand, he may change his beliefs, or he may suspend belief as to whether P is true. H may not believe $\neg P$, but simply believe that S is neiter competent nor sincere, and so may not come to believe P. The problem one is then faced with is this: How does one describe the effect of uttering the declarative sentence so that given the appropriate contextual elements, any one of these possibilities can follow from the description?

One possible approach to this problem would be to find some fundamental, context-independent effect of informing that is true *every* time a declarative sentence is uttered. If one's general theory of the world and of rational behavior were sufficiently strong and detailed, any of the consequences of

attitude revision would be derivable from the basic effect in combination with the elaborate theory of rationality. The initial efforts made along this path [3,5] entailed the axiomatization the effects of speech acts as producing in the hearer the belief that the speaker wants him to recognize the latter's intention to hold some other belief. The effects were characterized by nestings of Goal and Bel operators, as in

$$\mathrm{Bel}(H, \mathrm{Goal}(S, \mathrm{Bel}(H, P))).$$

If the right conditions for attitude revision obtained, the conclusion $\mathrm{Bel}(H, P)$ would follow from the above assumption.

This general approach proved inadequate because there is in fact *no* such statement about beliefs about goals about beliefs that is true in *every* performance of a speech act. It is possible to construct a counterexample contradicting any such effect that might be postulated. In addition, long and complicated chains of reasoning are required to derive the simplest, most basic consequences of an utterance in situations in which all of the "normal" conditions obtain — a consequence that runs counter to one's intuitive expectations.

Cohen and Levesque [4] developed a speech act theory in a monotonic modal logic that incorporates context-dependent preconditions in the axioms that state the effects of a speech act. Their approach overcomes the theoretical difficulties of earlier context-independent attempts; however, if one desires to apply their theory in a practical computational system for reasoning about speech acts, one is faced with serious difficulties. Some of the context-dependent conditions that determine the effects of a speech act, according to their theory, involve statements about what an agent does *not* believe, as well as what he does believe. This means that for conclusions about the effect of speech acts to follow from the theory, it must include an explicit representation of an agent's ignorance as well as of his knowledge, which in practice is difficult or even impossible to achieve.

A further complication arises from the type of reasoning necessary for adequate characterization of the attitude revision process. A theory based on monotonic reasoning can only distinguish between belief and lack thereof, whereas one based on *nonmonotonic* reasoning can distinguish between belief (or its absence) as a consequence of known facts, and belief that follows as a default because more specific information is absent. To the extent that such a distinction plays a role in the attitude revision process, it argues for a formalization with a nonmonotonic character.

Our research is therefore motivated by the following observations: (1) earlier work demonstrates convincingly that any adequate speech-act theory must relate the effects of a speech act to context-dependent preconditions; (2) these preconditions must depend on the ignorance as well as on the knowledge of the relevant agents; (3) any practical system for reasoning about ignorance must be based on nonmonotonic reasoning; (4) existing speech act theories based on nonmonotonic reasoning cannot account for the facts of attitude revision resulting from the performance of speech acts.

2 Perrault's Default Theory of Speech Acts

As an alternative to monotonic theories, Perrault has proposed a theory of speech acts, based on an extension of Reiter's default logic [11] extended to include default-rule schemata. We shall summarize Perrault's theory briefly as it relates to informing and belief. The notation $p \Rightarrow q$ is intended as an abbreviation of the default rule of inference,

$$\frac{p : \mathrm{M}q}{q}.$$

Default theories of this form are called *normal*. Every normal default theory has at least one *extension*, i.e., a mutually consistent set of sentences sanctioned by the theory.

The operator $B_{x,t}$ represents Agent x's beliefs at time t and is assumed to posess all the properties of the modal system weak $S5$ (that is, $S5$ without the schema $B_{x,t}\phi \supset \phi$), plus the following axioms:
Persistence:

$$B_{x,t+1} B_{x,t} P \supset B_{x,t+1} P \tag{1}$$

Memory:

$$B_{x,t} P \supset B_{x,t+1} B_{x,t} P \tag{2}$$

Observability:

$$\mathrm{Do}_{x,t}\alpha \wedge \mathrm{Do}_{y,t}(\mathrm{Obs}(\mathrm{Do}_{x,t}(\alpha))) \supset B_{y,t+1}\mathrm{Do}_{x,t}(\alpha) \quad (3)$$

Belief Transfer:

$$B_{x,t}B_{y,t}P \Rightarrow B_{x,t}P \quad (4)$$

Declarative:

$$\mathrm{Do}_{x,t}(\mathrm{Utter}(P)) \Rightarrow B_{x,t}P \quad (5)$$

In addition, there is a default-rule schema stating that, if $p \Rightarrow q$ is a default rule, then so is $B_{x,t}p \Rightarrow B_{x,t}q$ for any agent x and time t.

Perrault could demonstrate that, given his theory, there is an extension containing all of the desired conclusions regarding the beliefs of the speaker and hearer, starting from the fact that a speaker utters a declarative sentence and the hearer observes him uttering it. Furthermore, the theory can make correct predictions in cases in which the usual preconditions of the speech act do not obtain. For example, if the speaker is lying, but the hearer does not recognize the lie, then the hearer's beliefs are exactly the same as when the speaker tells the truth; moreover the speaker's beliefs about mutual belief are the same, but he still does not believe the proposition he uttered — that is, he fails to be convinced by his own lie.

3 Problems with Perrault's Theory

A serious problem arises with Perrault's theory concerning reasoning about an agent's ignorance. His theory predicts that a speaker can convince himself of any unsupported proposition simply by asserting it, which is clearly at odds with our intuitions. Suppose that it is true of speaker s that $\neg B_{s,t}P$. Suppose furthermore that, for whatever reason, s utters P. In the absence of any further information about the speaker's and hearer's beliefs, it is a consequence of axioms (1)–(5) that $B_{s,t+1}B_{h,t+1}P$. From this consequence and the belief transfer rule (4) it is possible to conclude $B_{s,t+1}P$. The strongest conclusion that can be derived about s's beliefs at $t+1$ without using this default rule is $B_{s,t+1}\neg B_{s,t}P$, which is not sufficient to override the default.

This problem does not admit of any simple fixes. One clearly does not want an axiom or default rule of the form that asserts what amounts to "ignorance persists" to defeat conclusions drawn from speech acts. In that case, one could never conclude that anyone ever learns anything as a result of a speech act. The alternative is to weaken the conditions under which the default rules can be defeated. However, by adopting this strategy we are giving up the advantage of using normal defaults. In general, nonnormal default theories do not necessarily have extensions, nor is there any proof procedure for such logics.

Perrault has intentionally left open the question of how a speech act theory should be integrated with a general theory of action and belief revision. He finesses this problem by introducing the persistence axiom, which states that beliefs always persist across changes in state. Clearly this is not true in general, because actions typically change our beliefs about what is true of the world. Even if one considers only speech acts, in some cases one can get an agent to change his beliefs by saying something, and in other cases not. Whether one can or not, however, depends on what belief revision strategy is adopted by the respective agents in a given situation. The problem cannot be solved by simply adding a few more axioms and default rules to the theory. Any theory that allows for the possibility of describing belief revision must of necessity confront the problem of inconsistent extensions. This means that, if a hearer initially believes $\neg p$, the default theory will have (at least) one extension for the case in which his belief that $\neg p$ persists, and one extension in which he changes his mind and believes p. Perhaps it will even have an extension in which he suspends belief as to whether p.

The source of the difficulties surrounding Perrault's theory is that the default logic he adopts is unable to describe the attitude revision that occurs in consequence of a speech act. It is not our purpose here to state what an agent's belief revision strategy should be. Rather we introduce a framework within which a variety of belief revision strategies can be accomodated efficiently, and we demonstrate that this framework can be applied in

a way that eliminates the problems with Perrault's theory.

Finally, there is a serious practical problem faced by anyone who wishes to implement Perrault's theory in a system that reasons about speech acts. There is no way the belief transfer rule can be used efficiently by a reasoning system; even if it is assumed that its application is restricted to the speaker and hearer, with no other agents in the domain involved. If it is used in a backward direction, it applies to its own result. Invoking the rule in a forward direction is also problematic, because in general one agent will have a very large number of beliefs (even an infinite number, if introspection is taken into account) about another agent's beliefs, most of which will be irrelevant to the problem at hand.

4 Hierarchic Autoepistemic Logic

Autoepistemic (AE) logic was developed by Moore [10] as a reconstruction of McDermott's nonmonotonic logic [9]. An autoepistemic logic is based on a first-order language augmented by a modal operator L, which is interpreted intuitively as self belief. A *stable expansion* (analogous to an extension of a default theory) of an autoepistemic base set A is a set of formulas T satisfying the following conditions:

1. T contains all the sentences of the base theory A
2. T is closed under first-order consequence
3. If $\phi \in T$, then $L\phi \in T$
4. If $\phi \notin T$, then $\neg L\phi \in T$

Hierarchic autoepistemic logic (HAEL) was developed in response to two deficiences of autoepistemic logic, when the latter is viewed as a logic for automated nonmonotonic reasoning. The first is a representational problem: how to incorporate preferences among default inferences in a natural way within the logic. Such preferences arise in many disparate settings in nonmonotonic reasoning — for example, in taxonomic hierarchies [6] or in reasoning about events over time [12]. To some extent, preferences among defaults can be encoded in AE logic by introducing auxiliary information into the statements of the defaults, but this method does not always accord satisfactorily with our intuitions. The most natural statement of preferences is with respect to the multiple *expansions* of a particular base set, that is, we prefer certain expansions because the defaults used in them have a higher priority than the ones used in alternative expansions.

The second problem is computational: how to tell whether a proposition is contained within the desired expansion of a base set. As can be seen from the above definition, a stable expansion of an autoepistemic theory is defined as a fixedpoint; the question of whether a formula belongs to this fixedpoint is not even semidecidable. This problem is shared by all of the most popular nonmonotonic logics. The usual recourse is to restrict the expressive power of the language, e.g., normal default theories [11] and separable circumscriptive theories [8]. However, as exemplified by the difficulties of Perrault's approach, it may not be easy or even possible to express the relevant facts with a restricted language.

Hierarchical autoepistemic logic is a modification of autoepistemic logic that addresses these two considerations. In HAEL, the primary structure is not a single uniform theory, but a collection of subtheories linked in a hierarchy. Subtheories represent different sources of information available to an agent, while the hierarchy expresses the way in which this information is combined. For example, in representing taxonomic defaults, more specific information would take precedence over general attributes. HAEL thus permits a natural expression of preferences among defaults. Furthermore, given the hierarchical nature of the subtheory relation, it is possible to give a constructive semantics for the autoepistemic operator, in contrast to the usual self-referential fixedpoints. We can then arrive easily at computational realizations of the logic.

The language of HAEL consists of a standard first-order language, augmented by a indexed set of unary modal operators L_i. If ϕ is any sentence (containing no free variables) of the first-order language, then $L_i\phi$ is also a sentence. Note that neither nesting of modal operators nor quantifying

into a modal context is allowed. Sentences without modal operators are called *ordinary*.

An HAEL structure τ consists of an indexed set of subtheories τ_i, together with a partial order on the set. We write $\tau_i \prec \tau_j$ if τ_i precedes τ_j in the order. Associated with every subtheory τ_i is a base set A_i, the initial sentences of the structure. Within A_i, the occurrence of L_j is restricted by the following condition:

$$\text{If } L_j \text{ occurs positively (negatively) in } A_i, \text{ then } \tau_j \preceq \tau_i \ (\tau_j \prec \tau_i). \tag{6}$$

This restriction prevents the modal operator from referring to subtheories that succeed it in the hierarchy, since $L_j\phi$ is intended to mean that ϕ is an element of the subtheory τ_j. The distinction between positive and negative occurrences is simply that a subtheory may represent (using L) which sentences it contains, but is forbidden from representing what it does *not* contain.

A *complex stable expansion* of an HAEL structure τ is a set of sets of sentences T_i corresponding to the subtheories of τ. It obeys the following conditions (ϕ is an ordinary sentence):

1. Each T_i contains A_i
2. Each T_i is closed under first-order consequence
3. If $\phi \in T_j$, and $\tau_j \preceq \tau_i$, then $L_j\phi \in T_i$
4. If $\phi \notin T_j$, and $\tau_j \prec \tau_i$, then $\neg L_j\phi \in T_i$
5. If $\phi \in T_j$, and $\tau_j \prec \tau_i$, then $\phi \in T_i$.

These conditions are similar to those for AE stable expansions. Note that, in (3) and (4), T_i contains modal atoms describing the contents of subtheories beneath it in the hierarchy. In addition, according to (5) it also inherits all the ordinary sentences of preceeding subtheories.

Unlike AE base sets, which may have more than one stable expansion, HAEL structures have a unique minimal complex stable expansion (see Konolige [7]). So we are justified in speaking of "the" theory of an HAEL structure and, from this point on, we shall identify the subtheory τ_i of a structure with the set of sentences in the complex stable expansion for that subtheory.

Here is a simple example, which can be interpreted as the standard "typically birds fly" default scenario by letting $F(x)$ be "x flies," $B(x)$ be "x is a bird," and $P(x)$ be "x is a penguin."

$$\begin{aligned} &\tau_0 \prec \tau_1 \prec \tau_2 \\ &A_0 = \{P(a), B(a)\} \\ &A_1 = \{L_1P(a) \wedge \neg L_0F(a) \supset \neg F(a)\} \\ &A_2 = \{L_2B(a) \wedge \neg L_1\neg F(a) \supset F(a)\} \end{aligned} \tag{7}$$

Theory τ_0 contains all of the first-order consequences of $P(a)$, $B(a)$, $L_0P(a)$, and $L_0B(a)$. $\neg L_0F(a)$ is *not* in τ_0, but it is in τ_1, as is $L_0P(a)$, $\neg L_0\neg P(a)$, etc. Note that $P(a)$ is inherited by τ_1; hence $L_1P(a)$ is in τ_1. Given this, by first-order closure $\neg F(a)$ is in τ_1 and, by inheritance, $L_1\neg F(a)$ is in τ_2, so that $F(a)$ cannot be derived there. On the other hand, τ_2 inherits $\neg F(a)$ from τ_1.

Note from this example that information present in the lowest subtheories of the hierarchy percolates to its top. More specific evidence, or preferred defaults, should be placed lower in the hierarchy, so that their effects will block the action of higher-placed evidence or defaults.

HAEL can be given a constructive semantics that is in accord with the closure conditions. When the inference procedure of each subtheory is decidable, an obvious decidable proof method for the logic exists. The details of this development are too complicated to be included here, but are described by Konolige [7]. For the rest of this paper, we shall use a propositional base language; the derivations can be readily checked.

5 A HAEL Theory of Speech Acts

We demonstrate here how to construct a hierarchic autoepistemic theory of speech acts. We assume that there is a hierarchy of autoepistemic subtheories as illustrated in Figure 1. The lowest subtheory, τ_0, contains the strongest evidence about the speaker's and hearer's mental states. For example, if it is known to the hearer that the speaker is lying, this information goes into τ_0.

In subtheory τ_1, defaults are collected about the effects of the speech act on the beliefs of both hearer and speaker. These defaults can be overridden by the particular evidence of τ_0. Together

τ_0 and τ_1 constitute the first level of reasoning about the speech act. At Level 2, the beliefs of the speaker and hearer that can be deduced in τ_1 are used as evidence to guide defaults about nested beliefs, that is, the speaker's beliefs about the hearer's beliefs, and vice versa. These results are collected in τ_2. In a similar manner, successive levels contain the result of one agent's reflection upon his and his interlocutor's beliefs and intentions at the next lower level. We shall discuss here how Levels τ_0 and τ_1 of the HAEL theory are axiomatized, and shall extend the axiomatization to the higher theories by means of axiom schemata.

An agent's belief revision strategy is represented by two features of the model. The position of the speech act theory in the general hierarchy of theories determines the way in which conclusions drawn in those theories can defeat conclusions that follow from speech acts. In our model, the speech act defaults will go into the subtheory τ_1, while evidence that will be used to defeat these defaults will go in τ_0. In addition, the axioms that relate τ_1 to τ_0 determine precisely what each agent is willing to accept from τ_0 as evidence against the default conclusions of the speech act theory.

It is easy to duplicate the details of Perrault's analysis within this framework. Theory τ_0 would contain all the agents' beliefs prior to the speech act, while the defaults of τ_1 would state that an agent believed the utterance P if he did not believe its negation in τ_0. As we have noted, this analysis does not allow for the situation in which the speaker utters P without believing either it or its opposite, and then becomes convinced of its truth by the very fact of having uttered it — nor does it allow the hearer to change his belief in $\neg P$ as a result of the utterance.

We choose a more complicated and realistic expression of belief revision. Specifically, we allow an agent to believe P (in τ_1) by virtue of the utterance of P only if he does not have any evidence (in τ_0) against believing it. Using this scheme, we can accommodate the hearer's change of belief, and show that the speaker is not convinced by his own efforts.

We now present the axioms of the HAEL theory for the declarative utterance of the proposition P. The language we use is a propositional modal one for the beliefs of the speaker and hearer. Agents s and h represent the speaker and hearer; the subscripts i and f represent the initial situation and the situation resulting from the utterance, respectively. There are two operators: $[a]$ for a's belief and $\{a\}$ for a's goals. The formula $[h_f]\phi$, for example, means that the hearer believes ϕ in the final situation, while $\{s_i\}\phi$ means that the speaker intended ϕ in the initial situation. In addition, we use a phantom agent u to represent the content of the utterance and certain assumptions about the speaker's intentions. We do not argue here as to what constitutes the correct logic of these operators; a convenient one is weak $S5$.

The following axioms are assumed to hold in all subtheories.

$$[u]P, \quad P \text{ the propositional content of utterance} \tag{8}$$

$$[u]\phi \supset [u]\{s_i\}[h_f]\phi \tag{9}$$

$$[a]\{a\}\phi \equiv \{a\}\phi, \quad \text{where } a \text{ is any agent in any situation.} \tag{10}$$

The contents of the u theory are essentially the same for all types of speech acts. The precise effects upon the speaker's and hearer's mental states is determined by the propositional content of the utterance and its mood. We assume here that the speaker utters a simple declarative sentence, (Axiom 8), although a similar analysis could be done for other types of sentences, given a suitable representation of their propositional content. Propositions that are true in u generally become believed by the speaker and hearer in τ_1, provided that these propositions bear the proper relationship to their beliefs in τ_0. Finally, the speaker *intends* to bring about each of the beliefs the hearer acquires in τ_1, also subject to the caveat that it is consistent with his beliefs in τ_0.

Relation between subtheories:

$$\tau_0 \prec \tau_1 \tag{11}$$

Speaker's beliefs as a consequence of the speech act:

$$\text{in } A_1: \quad [u]\phi \wedge \neg L_0 \neg [s_f]\phi \supset [s_f]\phi \tag{12}$$

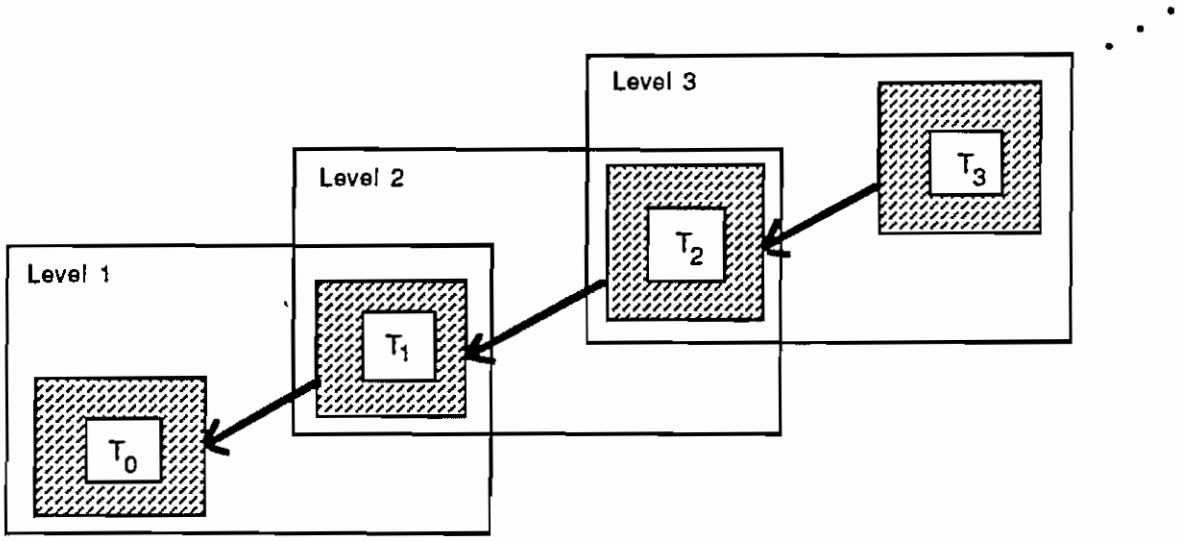

Figure 1: A Hierarchic Autoepistemic Theory

Hearer's beliefs as a consequence of the speech act:

$$\begin{aligned}&\text{in } A_1:\\ &\quad([u]\phi \wedge \\ &\quad\neg L_0\neg[h_J]\phi \wedge \neg L_0[h_J]\neg[s_J]\phi \wedge \\ &\quad\neg L_0[h_J]\neg\{s_i\}[h_J]\phi) \supset [h_J]\phi\end{aligned} \tag{13}$$

The asymmetry between Axioms 12 and 13 is a consequence of the fact that a speech act has different effects on the speaker's and hearer's mental states. The intuition behind these axioms is that a speech act should never change the speaker's mental attitudes with regard to the proposition he utters. If he utters a sentence, regardless of whether he is lying, or in any other way insincere, he should believe P after the utterance if and only if he believed it before. However, in the hearer's case, whether he believes P depends not only on his prior mental state with respect to P, but also on whether he believes that the speaker is being sincere. Axiom 13 states that a hearer is willing to believe what a speaker says if it does not conflict with his own beliefs in τ_0, and if the utterance does not conflict with what the hearer believes about the speaker's mental state, (i.e., that the speaker is not lying), and if he believes that believing P is consistent with his beliefs about the speaker's prior intentions (i.e., that the speaker is using the utterance with communicative intent, as distinct from, say, testing a microphone).

As a first example of the use of the theory, consider the normal case in which A_0 contains no evidence about the speaker's and hearer's beliefs after the speech act. In this event, A_0 is empty and A_1 contains Axioms 8–10. By the inheritance conditions, τ_1 contains $\neg L_0\neg[s_J]P$, and so must contain $[s_J]P$ by axiom 12. Similarly, from Axiom 13 it follows that $[h_J]P$ is in τ_1. Further derivations lead to $\{s_i\}[h_J]P$, $\{s_i\}[h_J]\{s_i\}[h_J]P$, and so on.

As a second example, consider the case in which the speaker utters P, perhaps to convince the hearer of it, but does not himself believe either P or its negation. In this case, τ_0 contains $\neg[s_J]P$ and $\neg[s_J]\neg P$, and τ_1 must contain $L_0\neg[s_J]P$ by the inheritance condition. Hence, the application of Axiom 12 will he blocked, and so we cannot conclude in τ_1 that the speaker believes P. On the other hand, since none of the antecedents of Axiom 13 are affected, the hearer does come to believe it.

Finally, consider belief revision on the part of the hearer. The precise path belief revision takes depends on the contents of τ_0. If we consider the hearer's belief to be stronger evidence than that of the utterance, we would transfer the hearer's initial belief $[h_i]\neg P$ to $[h_J]\neg P$ in τ_0, and block the default Axiom 13. But suppose the hearer does not believe $\neg P$ strongly in the initial situation. Then

we would transfer (by default) the belief $[h_f]\neg P$ to a subtheory higher than τ_1, since the evidence furnished by the utterance is meant to override the initial belief. Thus, by making the proper choices regarding the transfer of initial beliefs in various subtheories, it becomes possible to represent the revision of the hearer's beliefs.

This theory of speech acts has been presented with respect to declarative sentences and representative speech acts. To analyze imperative sentences and directive speech acts, it is clear in what direction one should proceed, although the required augmentation to the theory is quite complex. The change in the utterance theory that is brought about by an imperative sentence is the addition of the belief that the speaker intends the hearer to bring about the propositional content of the utterance. That would entail substituting the following effect for that stated by Axiom 8:

$$[u]\{s_f\}P, \quad P \text{ the propositional content of utterance} \tag{14}$$

One then needs to axiomatize a theory of *intention* revision as well as belief revision, which entails describing how agents adopt and abandon intentions, and how these intentions are related to their beliefs about one another. Cohen and Levesque have advanced an excellent proposal for such a theory [4], but any discussion of it is far beyond the scope of this article.

6 Reflecting on the Theory

When agents perform speech acts, not only are their beliefs about the uttered proposition affected, but also their beliefs about one another, to arbitrayr levels of reflection.

If a speaker reflects on what a hearer believes about the speaker's own beliefs, he takes into account not only the beliefs themselves, but also what he believes to be the hearer's belief revision strategy, which, according to our theory, is reflected in the hierarchical relationship among the theories. Therefore, reflection on the speech-act-understanding process takes place at higher levels of the hierarchy illustrated in Figure 1. For example, if Level 1 represents the speaker's reasoning about what the hearer believes, then Level 2 represents the speaker's reasoning about the hearer's beliefs about what the speaker believes.

In general, agents may have quite complicated theories about how other agents apply defaults. The simplest assumption we can make is that they reason in a uniform manner, exactly the same as the way we axiomatized Level 1. Therefore, we extend the analysis just presented to arbitrary reflection of agents on one another's belief by proposing axiom schemata for the speaker's and hearer's beliefs at each level, of which Axioms 12 and 13 are the Level 1 instances. We introduce a schematic operator $[(s,h)_n]$ which can be thought of as n levels of alternation of s's and h's beliefs about each other. This is stated more precisely as

$$[(s,h)_n]\phi \equiv_{def} \underbrace{\ldots[s][h]\ldots[s]}_{\text{n times}}\phi \tag{15}$$

Then, for example, Axiom 12 can be restated as the general schema

$$\begin{aligned} &\text{in } A_{n+1}\text{:} \\ &\quad ([u]\phi \wedge \\ &\quad \neg L_n[(h_f,s_f)_n]\neg[s_f]\phi) \supset \\ &\qquad [(h_f,s_f)_n][s_f]\phi. \end{aligned} \tag{16}$$

7 Conclusion

A theory of speech acts based on default reasoning is elegant and desirable. Unfortunately, the only existing proposal that explains how this should be done suffers from three serious problems: (1) the theory makes some incorrect predictions; (2) the theory cannot be integrated easily with a theory of action; (3) there seems to be no efficient implementation strategy. The problems are stem from the theory's formulation in normal default logic. We have demonstrated how these difficulties can be overcome by formulating the theory instead in a version of autoepistemic logic that is designed to combine reasoning about belief with autoepistemic reasoning. Such a logic makes it possible to formalize a description of the agents' belief revision processes that can capture observed facts about attitude revision correctly in response to speech acts. This theory has been tested and implemented as a central component of the GENESYS utterance-planning system.

Acknowledgements

This research was supported in part by a contract with the Nippon Telegraph and Telephone Corporation, in part by the Office of Naval Research under Contract N00014-85-C-0251, and in part under subcontract with Stanford University under Contract N00039-84-C-0211 with the Defense Advanced Research Projects Agency. The original draft of this paper has been substantially improved by comments from Phil Cohen, Shozo Naito, and Ray Perrault. The authors are also grateful to the participants in the Artificial Intelligence *Principia* seminar at Stanford for providing their stimulating discussion of these and related issues.

References

[1] Douglas E. Appelt. *Planning English Sentences.* Cambridge University Press, Cambridge, England, 1985.

[2] Philip R. Cohen. *On Knowning What to Say: Planning Speech Acts.* PhD thesis, University of Toronto, 1978.

[3] Philip R. Cohen and H. Levesque. Speech acts and rationality. In *Proceedings of the 23rd Annual Meeting*, pages 49–59, Association for Computational Linguistics, 1985.

[4] Philip R. Cohen and H. Levesque. *Rational Interaction as the Basis for Communication.* Technical Report, Center for the Study of Language and Information, 1987.

[5] Philip R. Cohen and C. Raymond Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:117–212, 1979.

[6] D. W. Etherington and R. Reiter. On inheritance hierarchies with exceptions. In *Proceedings of AAAI*, 1983.

[7] Kurt Konolige. *A Hierarchic Autoepistemic Logic.* Forthcoming technical note, 1988.

[8] Vladmir Lifschitz. Computing circumscription. In *Proceedings of AAAI*, pages 121–127, 1985.

[9] Drew McDermott. Nonmonotonic logic II: nonmonotonic modal theories. *Journal of the Association for Computing Machinery*, 29(1):33–57, 1982.

[10] Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1), 1985.

[11] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13, 1980.

[12] Yoàv Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, Cambridge, Massachusetss, 1987.